Using Social Media Content in the Visual Analysis of Movement Data

Robert Krüger **Steffen Lohmann** **Dennis Thom** **Harald Bosch** **Thomas Ertl**

Visualization and Interactive Systems Institute (VIS)
University of Stuttgart, Germany
{forename.lastname}@vis.uni-stuttgart.de

ABSTRACT

Data about the movement of people and objects is a rich source for visual analysis. However, understanding the data and inferring user behavior from it is often difficult due to missing context information. The goal of our research is to augment movement data by information derived from social media. In this paper, we present a visual concept that extends movement trajectories with terms extracted from geo-coded Twitter posts. The movement data comes from a large sample of e-bikes equipped with GPS devices. The Twitter terms are displayed as word clouds to provide additional context information for visual analysis. They are shown at locations of interests to help understand the movement data and infer possible user behavior. We plan to extend the visual concept in the future by incorporating further social media services. Ultimately, we aim for an integrated view of social media contents that allows for semi-automated reasoning and causality discovery on movement data.

Author Keywords

Movement data; social media; twitter; word cloud; visual analysis; movement trajectories; e-mobility; microposts.

INTRODUCTION

Data about the movement of people and objects is increasingly available with the widespread use of positioning techniques. More and more vehicles and mobile devices are equipped with GPS, allowing to track movement data of different type and granularity. Especially the aggregated movements of a large number of vehicles and/or people are a good source to better understand traffic flows and human behavior [1].

However, a well-known problem in analysis of movement data is the lack of context, i.e. we know *where* people go but we do not know *why* they go there. Though geo-spatial information about places, such as restaurants, theaters, or sports fields, is usually available, it often does not provide sufficient context to understand human behavior. Hence, the goal of our research is to help analysts in getting more insight from movement data by providing additional context information.

A potentially rich source for such context information is social media, which offers a vast amount of content from a multitude of people. In our case, especially geo-coded information associated with specific locations on a map is of interest. Some social media content is already used in web mapping services. For instance, Google Maps allows to display geotagged photos and videos from the media sharing websites Panoramio and YouTube on a map. Furthermore, it links ratings from review sites like TripAdvisor or Qype for many places.

However, there is a lot more information in social media that can support the visual analysis of movement data. In this work, we are particularly interested in text content, as posted in short messages to social networking and microblogging services like Twitter. Many of these *microposts* have a spatiotemporal character and are a useful complement to information available in other sources [2, 4].

While an individual post is small in size (up to 140 characters in Twitter) and usually of limited information value, aggregated posts of many people provide a rich source of time-sensitive information. From a technical perspective, users of microblogging services can be regarded as 'social sensors' [7]. They partly communicate local information that can help to understand movement data by pointing to events and other information not visible on a map.

The goal of the work presented in this paper is to augment movement data with content extracted from microblogging. We aim to integrate the content as smoothly as possible and provide the analyst with additional information about selected locations. In the following, we present a visual concept that illustrates our approach by augmenting movement data with terms extracted from Twitter.

COLLECTION AND VISUALIZATION OF MOVEMENT DATA

The movement data was collected in a large e-mobility project running from July 2010 to August 2011 in the Stuttgart metropolitan area. As part of the project, GPS devices were embedded in 500 e-bikes of type ELMOTO, which

To be presented at the '2nd Workshop on Interactive Visual Text Analytics'. This second edition of the workshop focuses on 'Task-Driven Analysis of Social Media Content' and will take place at VisWeek 2012 in Seattle, Washington, USA, on 15 October 2012.

can drive up to a speed of 45 kilometres per hour (see Figure 1). The bikes were given to a sample of people who could use them free of charge for the time of the project. In return, they agreed to the tracking of their movements. During their trips, the location of the e-bikes was recorded every 30 seconds, along with information about speed, battery charge, and other technical details. Overall, nearly one million kilometers of movement data were collected this way.

Figure 1. The movement data was obtained from 500 e-bikes equipped with GPS. © EnBW [`http://www.enbw.com/emobilitaet`]

To illustrate our visual concept, we took movement data from a one week period and plotted it on a map. We used a map from the OpenStreetMap project and converted it to grayscale. We then visualized the movement data as trajectories by connecting the recorded GPS locations with red lines. This nicely exposes the paths taken by the participants of the e-mobility project.

Figure 2 depicts a part of the map showing the largest campus of the University of Stuttgart. The highly aggregated visualization of the movement trajectories supports the identification of popular routes and spots. Of special interest are locations where a large number of trajectories end, i.e. where many people parked their bikes. Two such locations have been identified in the map section shown in Figure 2 (indicated by the green spots).

Apart from the names of some buildings, the map provides only few information that helps to understand the context of the selected locations. Further information would be useful to infer possible user behavior from the movement data. As discussed before, one source of such information can be microblogging.

CREATION OF WORD CLOUDS FROM TWITTER POSTS

Augmenting movement data with microblogging content is challenging due to the large volume and heterogeneity of publicly available microposts. They must usually be presented in highly aggregated form to be useful for visual analysis. Straightforward means to visually depict aggregated microposts are provided by *word clouds*. They proved useful in summarizing text content by displaying the most frequently used terms as a weighted list [5, 6]. We use this visualization technique to enrich locations of interest with terms extracted from microposts. As an illustration, we created word clouds for the two locations of interest shown in the map of Figure 2.

The microblogging content was retrieved from a large database of Twitter posts maintained by the VIS institute and described in [8]. The database contains most of the publicly available posts sent with geodata, which is nearly six million posts per day (as of August 2012). These posts are typically sent from mobile devices (e.g. smartphones). Roughly half of them contain a geotag (e.g. 'Stuttgart'), while the other half comes with precise latitude/longitude coordinates. We considered only the latter for the creation of the word clouds, as the former would be too imprecise to be plotted on a map with this level of detail.

Furthermore, we considered only microposts sent within a chosen time frame and originating from within a certain radius of the selected locations. The decision of what a good radius is depends on the characteristics of the location and the micropost density in that region. If we have a detailed map view of a well-populated area, as in our case, one would usually choose a relatively small radius in order to not let the context information interfere with that of a nearby location. The considered time frame would typically be the same as for the movement data, but it could also be longer or shorter, again depending on the interest of the analyst and the density of the available microposts.

The selected microposts are processed with techniques commonly used in word cloud generation, including tokenization, stop word removal, and conversion to lowercase [4, 5]. That is, we split the microposts into individual terms, converted these terms to lowercase, and removed common words that do not carry meaning. We then weighted each term based on how often it occurs in the microposts. Finally, we visualized the most frequently used terms in a circular word cloud. We chose this word cloud layout, because it supports the quick identification of the most popular terms [6, 8].

As illustrated in Figure 2, the word clouds provide additional context information that can help to better understand the movement data and related user behavior. For instance, the word cloud on the right hand site depicts the German term for university library ('universitätsbibliothek') in a large font size. This indicates that the library could be a reason why people parked their bikes at this location. As there is also a label for the library on the map, the additional context information would not have been necessary in this case. However, it further supports the aforementioned assumption.

In case of the left location of interest, by contrast, only few information is provided by the map alone. It shows neither names of buildings nor other information that helps in the interpretation of the movement data. Here, the terms extracted from microblogging and visualized in the word cloud provide at least some hints. For instance, the terms 'thesis' and 'learning' ('lernen') indicate that the buildings around the selected location also belong to the university. One interpretation would be that they are student hostels, which is further supported by terms like 'lunch' ('mittagessen') or 'roof terrace' ('dachterrasse'). However, additional context information, e.g. from other social media services, would be needed to get more evidence for this interpretation.

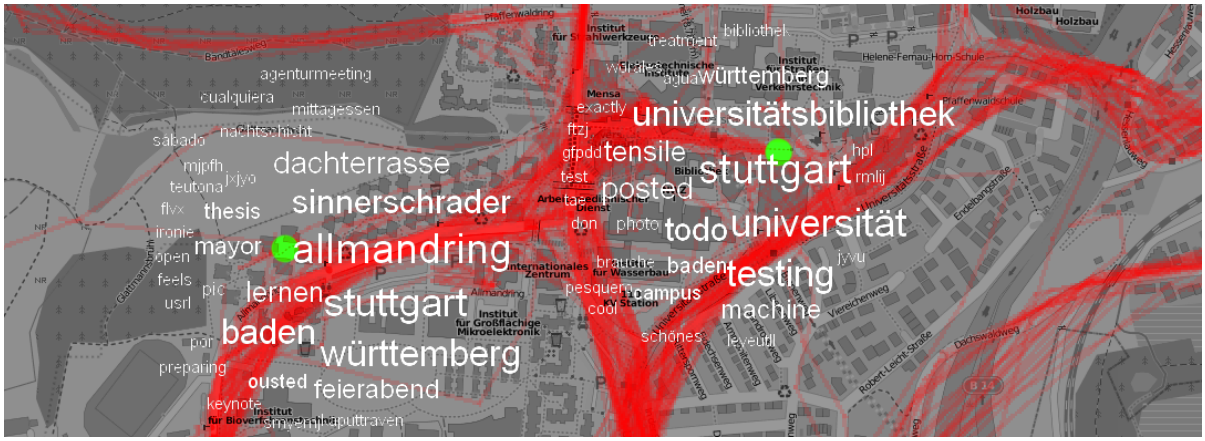

Figure 2. Map of movement trajectories augmented by terms extracted from geo-coded Twitter posts.

DISCUSSION AND FUTURE WORK

The presented approach gives an idea of the benefits that could be gained by integrating social media content in the visual analysis of movement data. The terms extracted from the microposts provide more information about locations of interest than is usually included in maps. They can thus contribute to a better understanding of the movement data and related user behavior.

On the other hand, the presented example illustrates the constraints and limitations of the approach. Most importantly, it depends on the microposts available for the selected time frame and location. While the number of microposts in the focused map area was sufficient to illustrate the concept, we expect better results for more densely populated areas where many people use Twitter on mobile devices. Therefore, we would like to test the visual concept also on movement data from other regions in the future.

Furthermore, we plan to advance the concept by implementing rich interactions, e.g. via filtering and highlighting. We also aim to aggregate movement trajectories of similar type into clusters, as proposed e.g. in [1], to ease the detection of locations of interest and recurring movement patterns. In addition, we are working on an improved relevance weighting of the terms extracted from Twitter. As only a small number of terms can be displayed in word clouds, it is important to include the most relevant ones. These are not necessarily the terms which occur most often in the microposts. Detecting, extracting, and weighting these terms is a non-trivial task that should at best take geographical context into account.

As we do not want to purely rely on Twitter as a source of context information, our overall goal is to also integrate content from other social media services, containing information about places, events, trends, etc. A big challenge will be a smooth and balanced integration of the contents without overwhelming the analyst with visual clutter. We plan to use integrated views based on automated reasoning and the 'intelligent' detection of relationships between the movement data and its context. Ideally, this leads to a visual analytics system supporting advanced tasks in movement analysis, such as causality discovery [3].

ACKNOWLEDGEMENTS

This work has been conducted in the contexts of the projects VASA (13N11254) and PESCaDO (FP7-248594) funded by the German Federal Ministry of Education and Research (BMBF) and the European Union. It was furthermore supported by the cooperative graduate program 'Digital Media' of the University of Stuttgart, the University of Tübingen, and the Stuttgart Media University (HdM).

REFERENCES

1. Andrienko, G., Andrienko, N., and Wrobel, S. Visual analytics tools for analysis of movement data. *SIGKDD Explor. Newsl. 9*, 2 (2007), 38–46.
2. Atkinson, C. *The backchannel: how audiences are using Twitter and social media and changing presentations forever*. New Riders, 2009.
3. Chen, M., Trefethen, A., Banares-Alcantara, R., Jirotka, M., Coecke, B., Ertl, T., and Schmidt, A. From data analysis and visualization to causality discovery. *IEEE Computer 44*, 10 (2011), 84–87.
4. Dörk, M., Gruen, D. M., Williamson, C., and Carpendale, M. S. T. A visual backchannel for large-scale events. *IEEE Trans. Vis. Comput. Graphics. 16*, 6 (2010), 1129–1138.
5. Gottron, T. Document word clouds: visualising web documents as tag clouds to aid users in relevance decisions. In *Proc. ECDL '09*, Springer (2009), 94–105.
6. Lohmann, S., Ziegler, J., and Tetzlaff, L. Comparison of tag cloud layouts: Task-related performance and visual exploration. In *Proc. INTERACT '09*, Springer (2009), 392–404.
7. Sakaki, T., Okazaki, M., and Matsuo, Y. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proc. of WWW '10*, ACM (2010), 851–860.
8. Thom, D., Bosch, H., Koch, S., Wörner, M., and Ertl, T. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *Proc. PacificVis '12*, IEEE (2012), 41–48.